A proposal for longer FITS keyword names

W. Pence, L. Chiappetti, M. Currie, J. Mink, R. Seaman

Version 0.68
26 January 2015

1 Introduction

The FITS scientific data format was invented in 1979 and has greatly benefited astronomical research by providing a common data format that is used by virtually all scientists and institutions in the field. FITS continues to serve astronomers well and will likely still be used for decades to come. It is widely acknowledged, however, that some of the original design decisions made when the format was created, while reasonable at the time, are now unnecessarily restrictive and inhibit new uses of FITS files in modern applications.

One of the most frequently cited restrictions is the 8-character limit on the length of keyword names in FITS headers. This restriction[1] negatively affects FITS file designers and users in several significant ways:

1. **Reduces Clarity** - The 8-character limit leads to cryptic mnemonic names that are confusing and difficult to remember for both software developers and end users of the data files. Examples can be found in almost every FITS file, such as `UCH2CJTM`, instead of a more descriptive name like `TEC_COLD_JUNCTION_2_TEMP`, or `ROTRTTRG`, instead of perhaps `TARGET_ROTATION_RATE`.

   This problem is exacerbated when dealing with arrays of indexed keywords or keywords associated with columns in a FITS table (e.g., TUNITn) where the root name is limited to even fewer characters to allow space for the numeric suffix. The World Coordinate System (WCS) keywords (as shown in Table 22 of Version 3 of the FITS standard) are a good illustration of this problem. In several instances, the root name of the WCS keyword is limited to just 2 characters (e.g., `12PC104A`) and in the most extreme case, the root name is reduced to a single letter (e.g., the letter `V` in `7V104_9A`) because of the need to also encode the column number, up to 2 coordinate axis numbers, a parameter sequence number, and the alternate WCS version code letter, all within only 8 characters!

2. **Hinders Innovation** - The keyword length limitation hinders innovation in developing new conventions for representing the complex data products that are being generated by current and future astronomical instruments. FITS binary tables, for example, offer great flexibility in storing complex data structures, but the 8-character limit is a major obstacle when trying to invent keyword names that also convey relationships or associations to other data elements (e.g., by adding an easily identifiable prefix to the names of the related set of keywords to define a unique 'namespace'). To cite another WCS example, in the draft paper on representations of distortions within coordinate systems, it was necessary to invent an entirely new type of 'record-valued' keyword structure because it is impossible to represent the large number of possible distortion keywords in any coherent way with only 8-characters names.

[1] While the 8-character limit is restrictive by today's standards, it was actually a significant increase over the 6-character limit on variable names in the Fortran-77 computer language which formed the basis of much of the FITS header record syntax.

3. **Hinders Data Exchange** - The keyword length restriction is also an obstacle when using FITS as an 'interchange format' for transporting data between different computer systems, which was the primary motivation for developing the FITS format in the first place. Other current scientific data formats allow meta-data parameters to have symbolic names longer than 8 characters, so representing them as FITS header keywords is problematic.

In order to help alleviate these problems, we propose a simple FITS convention for supporting longer FITS keyword names, as described in the following sections.

2 Proposed free-format convention for longer keyword names

The 8-character limit on the length of keyword names ultimately derives from the fact that the 'value indicator' (the equals sign character followed by a space character, '=␣', that separates the keyword name from the value field) is constrained to be in bytes 9 and 10 of the 80-byte keyword record. The obvious way to make room for longer keyword names is to lift this fixed format restriction and allow the value indicator to be located anywhere beyond bytes 9–10 of the keyword record.

The maximum keyword name length that can be supported by this proposed convention can be expressed as equal to 80-2-N, where N is the number of characters required in the value field (including the pair of enclosing quote characters in the case of a character string value). For example, keywords with a 16-bit integer value (with N = 6) could be up to 72 characters long, and keywords that express the full precision of a 64-bit floating point number in exponential notation (with N = 23) could be up to 55 characters long. Note that, in principle, the CONTINUE convention for encoding string values over multiple header records could be combined with this long keyword name convention to support arbitrarily long string values while also allowing keyword names up to 74 characters long.

Using keyword names that are close to the maximum possible length could be problematic, however, if the keyword needs to be updated with a new value that is longer than the available space in the keyword record. For this reason, it might be prudent to constrain the length of keyword names to, say, at most 55 characters, to ensure that enough space remains for all the most common types of numeric keyword values. We have tentatively decided not to place such a constraint on the name length, however, we believe that it would be useful to poll the larger FITS community on this matter before making a final implementation decision.

2.1 Side note on the legality of this FITS convention

One notable feature of this 'free-format' convention for supporting long keyword names is that it conforms to the current FITS format requirements for header records (as defined in Section 4.1 FITS standard) and thus, strictly speaking, requires no modifications to that document to be considered 'legal' FITS usage. Of particular relevance is sub-section 4.1.2.3, which is quoted here:

> In keyword records that contain the value indicator in bytes 9 and 10, the remaining bytes 11 through 80 of the record shall contain the value, if any, of the keyword, followed by optional comments. In keyword records without a value indicator, bytes 9 through 80 should be interpreted as commentary text, however, this does not preclude conventions that interpret the content of these bytes in other ways.

In other words, the default interpretation of keyword records that do not have the value indicator in bytes 9 and 10 is to treat the characters in bytes 1–8 as the keyword name and all the remaining characters in bytes 9–80 as commentary text. However, the standard explicitly allows other conventions to be established which interpret the keyword record differently.

In fact, this proposed convention is not the first to make use of this provision in the FITS standard: The ESO HIERARCH convention, with keywords of the form
`HIERARCH ESO INS OPTI-3 ID = 'ESO427 ' / Opt.  ID`
has been widely used since 1990 and is another example of a convention that ascribes an alternate interpretation to FITS header records that do not have a value indicator in bytes 9–10.

To be clear, although it is not strictly necessary to modify the FITS standard to allow the type of free-format keyword records defined by this convention, we do recommend that the discussion in section 4.1 of the standard be expanded to document and describe this new convention.

3 Optional enhancement: allow additional characters in the name

By default, this convention requires that the longer keyword names conform to the same restrictions as normal 8-character keyword names. The only characters that are allowed in the name are the Latin alphabetic characters `'A'` through `'Z'`, the digits 0 through 9, and the underscore and hyphen characters.

This may be an opportune time, however, to revisit this restriction and consider modifying the FITS standard to allow additional characters to be used within keyword names. In particular, the following characters might be especially useful:

- The 26 lower case letters `'a'`–`'z'`. Use of these character could make the names more legible to human readers. If lower case characters are allowed, then it also must be decided whether or not the case is significant (e.g., should `VOLTAGE_MAX`, `VOLTAGE_Max`, and `VOLTAGE_max` be interpreted as different keywords or not). The standard currently states that table column names *should not* be case sensitive, so perhaps the same interpretation should apply to keyword names, especially since in some situations column names and keywords are interchangeable (e.g., under the Green Bank Convention),

- Characters that might be useful as delimiters within a hierarchical keyword name, such as `'.'` (period), `':'` (colon), or `'$'` (dollar sign). These characters could be used to create keywords such as `HST:WFPC:VOLTAGE:MAXIMUM`.

- It also might be convenient to allow embedded space characters within keyword names, Doing so would mean that the ESO HIERARCH keywords would be naturally supported under this convention and would not have to be treated as a special case. Allowing spaces within keyword names could, however, cause problems for existing FITS software which might assume that the space character serves as a terminator of the keyword name. If the name includes spaces, then it also might require that the name be enclosed within quote characters to avoid any ambiguity about where the name ends.

4 `LONGKYWD` keyword

When using this convention, it is strongly recommended that the `LONGKYWD` keyword, with a value of 1.0, be added to the header of any HDU that contains (or may contain) keywords with long names. The presence of this `LONGKYWD` keyword serves to indicate that header records may use this

convention.[2] The value of 1.0 indicates that version 1 of this convention, as defined here, is used. A larger value could be defined in the future if a new version of this convention is developed.

5 Impact of this new convention on software

It is anticipated that use of this basic convention (i.e.,without expanding the allowed character set in keyword names) within new FITS files will have minimal impact on existing software packages. If legacy software does not support this convention, then according to the FITS standard, it should simply interpret the first 8 characters of the keyword record as the keyword 'name', and because there is no value indicator in bytes 9–10, the rest of the keyword record should be interpreted as commentary text, similar to the way `COMMENT` or `HISTORY` keywords are treated. Such legacy software will not be able to correctly interpret the intended name and value of this new type of keyword, of course, but as long as FITS file designers are judicious in the use of this convention (e.g., at least initially, only use this convention for new mission-specific keywords that are unlikely to be required by legacy software packages) this is not likely to be a significant problem. This is quite similar to the case of the ESO HIERARCH keywords which have been widely used for decades without any serious software compatibility issues.

If an expanded set of characters are allowed in keyword names, as discussed in Section 3, then this could have a larger impact and potentially could cause some applications software to be unable to process the FITS file until the software is updated to support the newly allowed characters in the name. For example, some FITS libraries (notably, CFITSIO) verify that the keyword name conforms to the current FITS requirements before writing or copying the keyword to a new FITS file. Thus, simply attempting to copy keywords from one FITS file to another file may trigger an error if the software finds what it considers to be an illegal character within the first 8 characters of the keyword record. In this case, the application program would need to be relinked with an updated version of the FITS library that supports the new characters to resolve this problem.

The cost of retrofitting existing software to fully support this convention is anticipated to be relatively modest. The biggest impact will be on the standard FITS libraries that most applications software packages rely on to read and write FITS keywords. As an example, a couple days of programming effort was required to upgrade the CFITSIO library to support a beta version of this convention .

Once a FITS library has been upgraded, the applications programs that are linked to that library will then inherit the ability to read and write longer keyword names. In some cases it may also be necessary to make further small modifications to the application program itself, for example, to increase the size of string variables within the code in order to store the longer keyword names in computer memory.

6 Summary

The 8-character keyword name limitation has a significant negative impact on current FITS users and is impeding the development of new FITS conventions. Removing this limitation will provide significant benefits to future projects by allowing them to create clearer, more self-documenting keyword names at a relatively small cost to existing software systems.

[2]Alternatively, one might consider defining a more global FITS version keyword that would attempt to convey all the different FITS format features and FITS conventions that are used within that particular FITS file or HDU. Defining such a keyword would require extensive discussion, however, and is considered to be beyond the scope of the present work.